

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

ADVANCED AND ROBUST MOBILE SEARCH ENGINE FOR PERSONALIZED

Mr. Yogesh B.Jadhao^{*1} and Assoc.Prof. Shyam S. Gupta ²

^{*1}Research scholar, Computer Engineering, Siddhant College of Engineering, Sadumbare, Pune, India
²Assoc.Prof. Shyam S. Gupta, Computer Engineering, Siddhant College of Engineering, Sadumbare, Pune, India

ABSTRACT

Data mining is becoming more popular. The main function of data mining is to retrieve hidden useful information from database. This paper uses data mining concept for collecting user's multiple response from click through data given by user. The collecting user feedback is based on the search results and the location of the user. PMSE describes privacy of the user and returns the result of user query. Here methodologies are introduced to balance the weights between the content and location facets which can able to gives exact result from different query data

Keywords: Click through data, concept, location search, mobile search engine, personalization.

I. INTRODUCTION

The technology is covering whole world. Everyone uses the mobile phone for browsing or searching on internet. Mobile search engine (MSE) is same as that of traditional search engine. It accepts the user query and produces the relevant result to the user. Now a day's people starts to use mobile phone for their day by day use like official, educational, family or other normal life purposes. And everyone wants all the information in one click. The Smartphone's are having capability to give enormous access of internet to the user. Due to this user can access information of topic at any place and any time at which he demands.

But sometimes a user wants to search a particular query on mobile search engine. And mobile search engine gives ambiguous results and showing the large list of irrelevant results which are not useful to the user. Due to this user get caught inside the loop of non-useful results. Many times user relevant results are not fetched by database then also the list of result is shown by the search engine.

Mobile device has small factor which limits the interaction between user and mobile device. To return the exact or user query relevant results the user profile must be used along with results. Profiling of user interest and personalized data will help to produce the highly relevant search result.

To produce more relevancy in the search results, we are going to use the location of the user. Along with that the visited physical locations of users in the MSE are also being considered. For accessing user location information mobile GPS or other location based services are used.

The GPS is location based service which is connected to the location server. This server returns all information about user location and user requested location.

For example if the user wants to search the information about 'Tigers'. And he is right now in 'Bengal', then Mobile search engine should return information about Bengal tigers. Then according to his location the search engine will give the information about Bengal Tigers. Because user is right now in Bengal and the information of other regions Tiger is not useful for him. Search engine will also consider the user profile while returning the results.

II. LITERATURE SURVEY

There are so many researchers have been done on the personal search mobile search engine. Their literature is stated as follow

A. Concept Based Clustering:

It consist of following steps. At starting when a user submits a query, then its relations are retrieved online from web-snippets to build a concept relationship graph. Next, clickthrough data are collected to decide user's preferences.

After that, the concept relationship graph and the user’s conceptual preferences is used to gather as input to a concept-based clustering algorithm that finds close queries and their results.

The idea of this method is based on content and their relations retrieved from the submitted user queries, the web-snippets and the clickthrough data. Clickthrough data was exploited in the personalized clustering process to identify user preferences: The user’s interest is computed depending upon collecting clicked snippet by user. User will click on that snippet in which he is interested. Clickthrough data can be stored easily without imposing extra burden on users. Concept based clustering goes in depth for finding out the synonyms or actual meaning of the query typed.

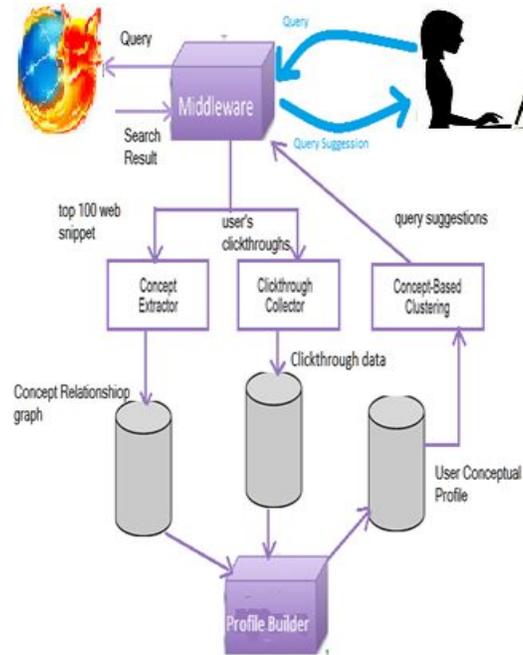


Figure 1. Architecture of concept based clustering.

A. Kokomo Search:

This is used to provide local information from the Internet by collecting, structuring, organizing, and filtering in a practicable form. To utilize such local information, Mobile Info Search employs a mediator architecture; a software agent between users and the information sources. The architecture is shown in Figure 2.

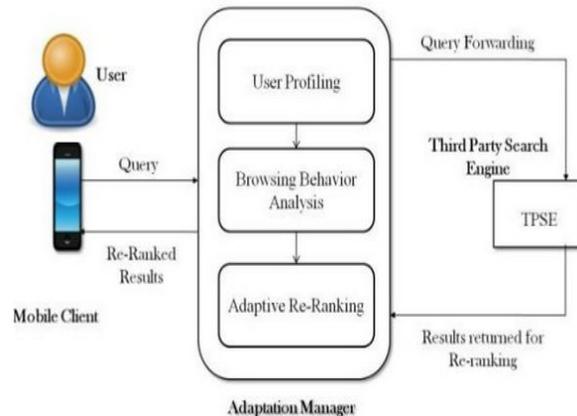


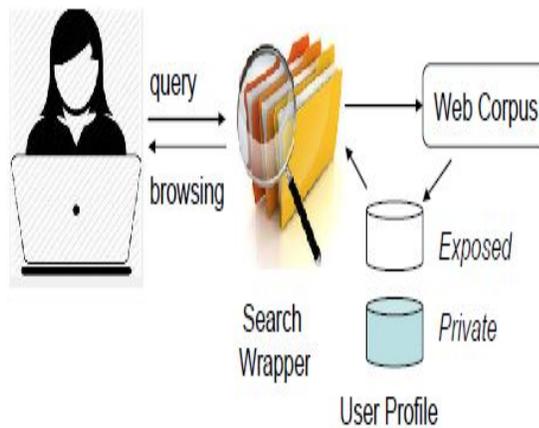
Figure 2. Kokono search

Between users and information sources, Mobile Info Search mediates Database-type resources using Location-Oriented Meta Search and static files using Location-Oriented Robot-based Search using Automatic Location Metadata Generation. To use the service, only Internet connectable PDSs or PCs with Web browsers are required for users.

C. Privacy Search:

User profile consists of personal documents, browsing history and emails which might be helpful to identify a user’s interests. But, users have knowledge about how their personal information is used. Privacy, as exactly oppose to security or confidentiality, highly depends on the person involved and how that person may benefit from sharing personal information. If user wants to share their profile then how to preserve user’s privacy is a big question. This algorithm is provided for the user to automatically build a generalised structural user profile that represents the user’s personal interests.

User’s interests are put on a top level; specific interests are put on a bottom level. According to users privacy setting the users profile will be shared to the server. It gives solution in three different parts: First, a scalable algorithm for automatically building a hierarchical user profile from available data. Second, privacy parameters are used to determine the content and amount of personal information that will be revealed. Third, search results are



displayed by search engine wrapper.

Figure:3. Privacy Search

D. Partially Supervised Classification:

This paper gives the description about following problems: From a set of documents of a class P, and a large set result that contains results from class P and remaining types of results, identify the results from class P in M. There is no labelled non- P document, which makes traditional machine learning techniques inapplicable, as they all need labelled documents of both classes. We call this problem partially supervised classification. Here this problem is called as partially supervised classification, as there are only positive results but not labelled negative results.

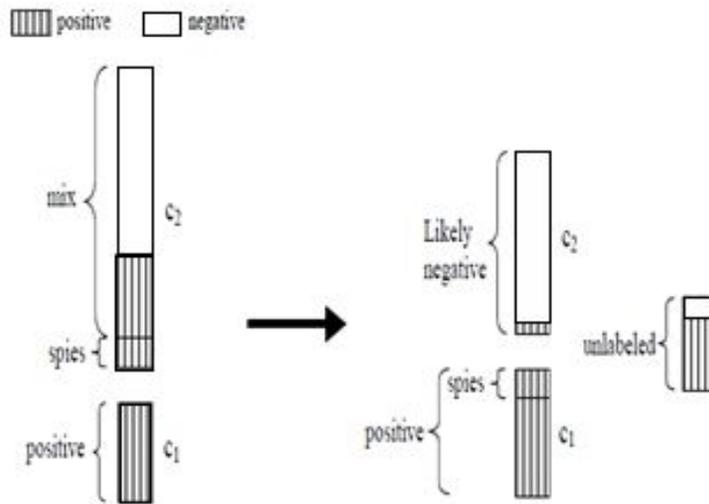


Figure: 4. Partially supervise classification.

III. PREPOSED SYSTEM

In proposed work we are going to use concept of ontology. Here we are combining the click through data and location of the user to predict the result. User mobile phone must be capable of using the location based system such as GPS. These services will return the User location to the search engine. Domain knowledge is used along with the user profile to produce most relevant results. We are providing the data to user beyond his expectations using his own profile so it is called as personalize search.

In this system client is responsible for storing the user click through and the ontologies accepted from the server. The Extractor handles other simple tasks such as updating creating feature extractor, obtaining reranked search results and clickthroughs and ontologies. The re-ranker handles other remaining tasks which may include reranking of search results. The user query is firstly hit by the feature extractor server then it will produce ranked results depending upon the user profile stored inside the feature extractor

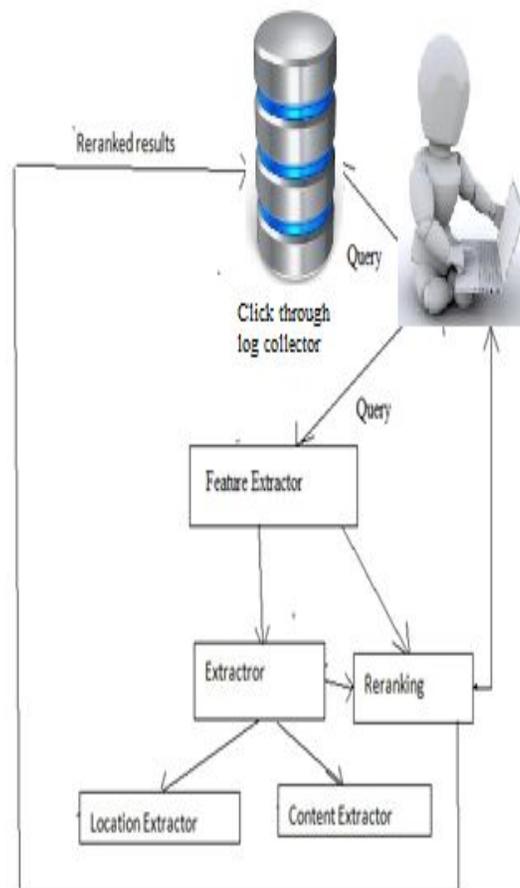


Figure:5. System Architecture

A. User Profiling.

User profile is nothing but the combination of two parameters. First one is content of the query result and second one is location of the user. We are using the ontology to determine the content profile. Click through collector and feature extractor combine to form a content profile. Second parameter is user location. Which is return by GPS or other Location based services integrated inside the mobile. The user entered query having some domain knowledge which is known ontology.

IV. RESULT & ANALYSIS

Initially we are accepting user query as input and then search content on the basis of location of user.

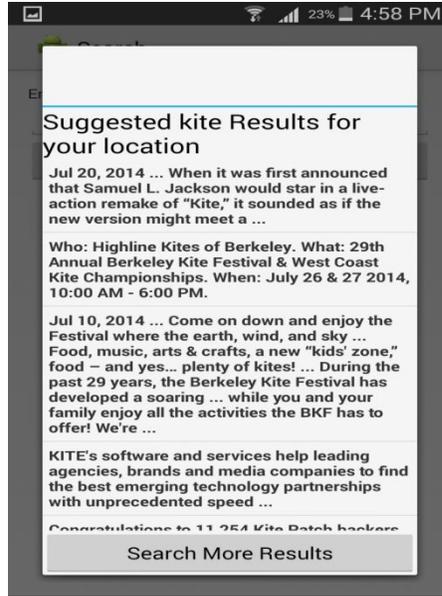


Figure 6. Location wise results.

After location based searching query results selected by user are added to the database as a part of history.

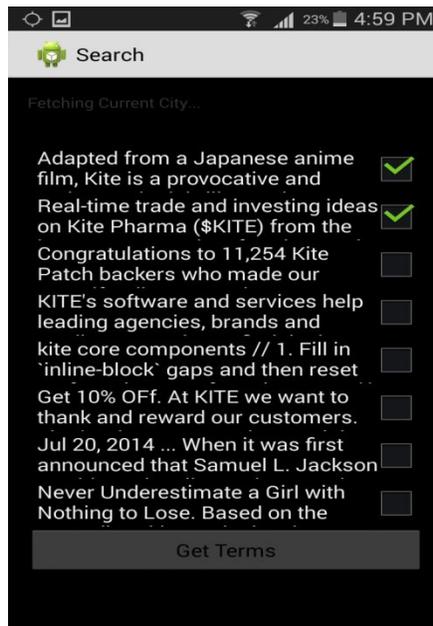


Figure: 7. Search result GUI for user interest

V. EXPECTED RESULTS

This system aims to find relevant results on the basis of several factors like history, domain knowledge, location and ontology. System should return most relevant result depending upon user interests. Ontology must extract accurate

and efficient result by cooperating with click through log collector. After result mining system should re-rank those results in order of user interests.

VI. CONCLUSION

Here we can conclude that our proposed system is a combination of user location, domain knowledge, history, clickthrough data and user profile. This paper uses the concept of ontology extract the content of the query. Location based ontology will generate more accurate results from the bunch of duplicate results. As a result of combination of all these methods, this system will produce the efficient output.

VII. ACKNOWLEDGEMENTS

We would like to thanks to Principal, Siddhant College of Engineering Sadumbare. We are also thankful to our family and friends for their encouragement and support.

REFERENCES

1. Kenneth Wai-Ting Leung, Dik Lun Lee, and Wang Chien Lee "PMSE: A Personalized Mobile Search Engine" *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, VOL. 25, NO. 4, APRIL 2013.
2. Nat'l geospatial, <http://earth-info.nga.mil/>, 2012.
3. svmlight, <http://svmlight.joachims.org/>, 2012.
4. World gazetteer, <http://www.world-gazetteer.com/>, 2012.
5. Agichtein, E. Brill, and S. Dumais, "Improving Web Search Ranking by Incorporating User Behavior Information," *Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR)*, 2006.
6. Agichtein, E. Brill, S. Dumais, and R. Ragno, "Learning User Interaction Models for Predicting Web Search Result Preferences," *Proc. Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR)*, 2006.
7. Y.-Y. Chen, T. Suel, and A. Markowetz, "Efficient Query Processing in Geographic Web Search Engines," *Proc. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR)*, 2006.
8. K.W. Church, W. Gale, P. Hanks, and D. Hindle, "Using Statistics in Lexical Analysis," *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*, Psychology Press, 1991.
9. Q. Gan, J. Attenberg, A. Markowetz, and T. Suel, "Analysis of Geographic Queries in a Search Engine Log," *Proc. First Int'l Workshop Location and the Web (LocWeb)*, 2008.
10. T. Joachims, "Optimizing Search Engines Using Clickthrough Data," *Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining*, 2002.
11. K.W.-T. Leung, D.L. Lee, and W.-C. Lee, "Personalized Web Search with Location Preferences," *Proc. IEEE Int'l Conf. Data Mining (ICDE)*, 2010.
12. K.W.-T. Leung, W. Ng, and D.L. Lee, "Personalized Concept-Based Clustering of Search Engine Queries," *IEEE Trans. Knowledge and Data Eng.*, vol. 20, no. 11, pp. 1505-1518, Nov. 2008.
13. Li, Z. Li, W.-C. Lee, and D.L. Lee, "A Probabilistic Topic-Based Ranking Framework for Location-Sensitive Domain Information Retrieval," *Proc. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR)*, 2009.
14. Liu, W.S. Lee, P.S. Yu, and X. Li, "Partially Supervised Classification of Text Documents," *Proc. Int'l Conf. Machine Learning (ICML)*, 2002.
15. W. Ng, L. Deng, and D.L. Lee, "Mining User Preference Using Spy Voting for Search Engine Personalization," *ACM Trans. Internet Technology*, vol. 7, no. 4, article 19, 2007.
16. J.Y.-H. Pong, R.C.-W. Kwok, R.Y.-K. Lau, J.-X. Hao, and P.C.-C. Wong, "A Comparative Study of Two Automatic Document Classification Methods in a Library Setting," *J. Information Science*, vol. 34, no. 2, pp. 213-230, 2008.

17. C.E. Shannon, "Prediction and Entropy of Printed English," *Bell Systems Technical J.*, vol. 30, pp. 50-64, 1951.
18. Q. Tan, X. Chai, W. Ng, and D. Lee, "Applying Co-Training to Clickthrough Data for Search Engine Adaptation," *Proc. Int'l Conf. Database Systems for Advanced Applications (DASFAA)*, 2004.
19. J. Teevan, M.R. Morris, and S. Bush, "Discovering and Using Groups to Improve Personalized Search," *Proc. ACM Int'l Conf. Web Search and Data Mining (WSDM)*, 2009.
20. E. Voorhees and D. Harman, *TREC Experiment and Evaluation in Information Retrieval*. MIT Press, 2005.
21. Y. Xu, K. Wang, B. Zhang, and Z. Chen, "Privacy-Enhancing Personalized Web Search," *Proc. Int'l Conf. World Wide Web (WWW)*, 2007.
22. S. Yokoji, "Kokono Search: A Location Based Search Engine," *Proc. Int'l Conf. World Wide Web (WWW)*, 2001.